

ABSTRACT OF THE DISCLOSURE

A method for and data structure for compressing file-reference data for source code identifiers compresses global reference data and local reference data. Compression of local data uses various methods to encode location data (e.g. line numbers and, optionally, column numbers) representing references to a source code information symbol in a source code file. Run length encoding and other run encoding methods are used. A preferred run encoding method encodes a run-encoded bitmap comprising a set of runs, in which each run has a first run field that provides a starting line for the run, and a second run field that encodes as a bitmap the remaining lines in the run in which the reference appears. A run length field may be used to indicate the length of the second run field. Another run field may be used to encode the length of the first run field, to permit further compression. The advantage is significant storage requirement reduction without sacrifice of information retrieval capacity.